COMS-4771 Spring 2008
Paper 3 Quiz Name: ________________________________

Avrim Blum and Tom Mitchell. *Combining Labeled and Unlabeled Data with Co-Training.*

You have 30 minutes to complete the questions. The quiz is worth 10 points.

Question 1 (3 points): Consider a joint probability distribution D between X_1 and X_2 (the two views), together with the outcome Y.

X_1	X_2	$D(X_1, X_2)$	Y
0	0	1/9	0
0	1	1/9	1
0	2	1/9	0
1	0	1/9	1
2	0	1/9	0
1	1	1/9	0
1	2	1/9	1
2	1	1/9	1
2	2	1/9	0

Does this distribution satisfy the assumptions of co-training?

Answer: No, X_1 and X_2 are not conditionally independent. The label Y is also not deterministic (thus the views are not always consistent) while co-training assumes that there is no noise.

Question 2 (7 points): Let $X = X_1 \times X_2$ be the instance space. Let $X^+ = X_1^+ \times X_2^+$ denote the positive region of X (the region where the label is 1), and let D^+ denote the marginal distribution of D over X^+. (Assume that the learning algorithm on each view can PAC-learn from positive examples only.)

For $S_1 \subseteq X_1^+$, define $p_1 = \mathbf{Pr}_{x_1 \sim D_1^+}[x_1 \in S_1]$, where D_1^+ is D^+ restricted to X_1^+. (Define p_2 similarly for $S_2 \subseteq X_2^+$.) Also define $p_{2|1} = \mathbf{Pr}_{(x_1,x_2)\sim D^+}[x_2 \in S_2 \mid x_1 \in S_1]$.

We say that D^+ is ϵ-*expanding* if

$$p_1 \leq 1/2 \quad \text{and} \quad p_{2|1} \geq 1 - \epsilon$$

together imply that $p_1 \geq (1+\epsilon)p_2$, for any $S_1 \subseteq X_1^+$, $S_2 \subseteq X_2^+$. (Similarly with views 1 and 2 reversed.)

Questions:

- Is the assumption that D^+ is ϵ-expanding stronger than the assumption of conditional independence? What does the conditional independence assumption imply in terms of expansion?
- Intuitively, why is expansion helpful in co-training?

Answer: No, the conditional independence assumption is stronger. Conditional independence implies that for any S_1 and S_2, we have $p_{2|1} = p_2$. So if $p_{2|1} \geq 1 - \epsilon$, then $p_2 \geq 1 - \epsilon$ as well, regardless of how small p_1 is. Conditional independence implies that not only does our confident set S_1 expand by a $(1 + \epsilon)$ factor as in the definition of ϵ-expansion, but it expands to nearly all of X_2^+.

Why is expansion useful? If S_1 is small ($p_1 \leq 1/2$) and we drive down the error on the conditional distribution induced by S_1 on X_2, the definition of expansion implies that the confident set on X_2 will have a noticeably larger probability than S_1.

There was an (unintentional) typo in the definition of expansion: the definion had $p_1 \geq (1-\epsilon)p_2$ instead of $p_1 \geq (1+\epsilon)p_2$ (so it's rather ϵ-shrinking not expanding). Surprisingly, no one noticed.

The notion of expansion was defined in *Co-training and Expansion*, by Balcan, Blum and Yang.